

University of Dundee

## Implementation fidelity and pupil achievement in book reading

Topping, Keith

*Published in:*  
Research Papers in Education

*DOI:*  
[10.1080/02671522.2017.1329340](https://doi.org/10.1080/02671522.2017.1329340)

*Publication date:*  
2018

*Document Version*  
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

### *Citation for published version (APA):*

Topping, K. (2018). Implementation fidelity and pupil achievement in book reading: variation between regions, local authorities and schools. *Research Papers in Education*, 33(5), 620-641.  
<https://doi.org/10.1080/02671522.2017.1329340>

### **General rights**

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Implementation Fidelity and Pupil Achievement in Book Reading:  
Variation between Regions, Local Authorities and Schools

Keith Topping

University of Dundee, Scotland

Implementation fidelity or integrity is a key variable in evidence-based interventions, but is rarely assessed by student response or over a long period. This study related implementation fidelity in independent book reading to tested reading achievement over a year. The sample of 852,295 students in 3243 primary and secondary schools was reduced by schools providing incomplete data and the discarding of mid-year data. Achievement was measured pre-post by STAR Reading, a computerised item-banked adaptive test of reading accuracy and comprehension. Implementation fidelity was measured by variables from the computerised Accelerated Reader (AR) software, which analyses comprehension of a real book the student has chosen by a quiz. Results compared key variables for the four regions of the UK, local authorities and best performing schools. Computerised methods offer an interesting alternative to teacher behaviour for investigating implementation and outcomes. This study showed stronger links between implementation and outcomes at an individual level than when all data is aggregated, and that student response is at least as good an implementation index as teacher behaviour. The implications for practice, policy and future research were outlined.

182 words

Running Head: VARIATION IN BOOK READING IMPLEMENTATION FIDELITY

Length: 8446 words excluding tables (a further 2135 words) (31 pages excluding tables, plus 14 pages tables)

Conflict of Interest: Dr. Topping is a Professor at the University of Dundee. The results of this study do not create a conflict of interest for him.

Financial Disclosure: The author has no relevant financial relationships to disclose.

Correspondence to: Professor Keith Topping, School of Education, University of Dundee, Dundee DD1 4HN, United Kingdom. Tel: +44 07854 833556, Email: [k.j.topping@dundee.ac.uk](mailto:k.j.topping@dundee.ac.uk)

N.B. In the schools analysis, permission has been obtained for the top three schools in each sector in each region to be named, which is done here. Permission can be sought from the remaining schools if the Editor so wishes.

### Abstract

Implementation fidelity or integrity is a key variable in evidence-based interventions, but is rarely assessed by student response or over a long period. This study related implementation fidelity in independent book reading to tested reading achievement over a year. The sample of 852,295 students in 3243 primary and secondary schools was reduced by schools providing incomplete data and the discarding of mid-year data. Achievement was measured pre-post by STAR Reading, a computerised item-banked adaptive test of reading accuracy and comprehension. Implementation fidelity was measured by variables from the computerised Accelerated Reader (AR) software, which analyses comprehension of a real book the student has chosen by a quiz. Results compared key variables for the four regions of the UK, local authorities and best performing schools. Computerised methods offer an interesting alternative to teacher behaviour for investigating implementation and outcomes. This study showed stronger links between implementation and outcomes at an individual level than when all data is aggregated, and that student response is at least as good an implementation index as teacher behaviour. The implications for practice, policy and future research were outlined.

Keywords: implementation fidelity, implementation integrity, book reading, reading achievement, United Kingdom, regions, local authorities, primary schools, secondary schools

## Implementation Fidelity and Pupil Achievement in Book Reading:

### Variation between Regions, Local Authorities and Schools

Keith Topping

Implementation fidelity has been defined as the degree to which an intervention or treatment is implemented as planned, intended, or originally designed (Lane, Bocian, MacMillan, & Gresham, 2004). Schulte, Easton and Parker (2009) espoused five elements in implementation fidelity often found in the previous literature: adherence to an intervention, exposure or dose, quality of delivery, participant responsiveness and programme differentiation (the extent to which key factors in effectiveness are identified). They also considered how the participant was able to use the learned skills in the natural environment. Carroll, Patterson, Wood, Booth, Rick, and Balain (2007) added two further elements: intervention complexity and facilitation strategies (i.e. strategies deliberately intended to enhance implementation quality, such as the provision of manuals, guidelines, training, monitoring and feedback, capacity building, and incentives). Of course, the question then arises of which of these many indices are most related to outcome (Durlak & DuPre, 2008).

Despite the importance of treatment fidelity, it has historically been largely overlooked. Since the emphasis has moved towards “evidence-based” methods and interventions, measuring the quality of implementation of an intervention has become an increasing preoccupation. Clearly, there is little point attempting to implement an evidence-based method and measure the outcomes if there is no parallel attempt to see whether the method has actually been implemented. As Carroll, et al. (2007) express it, implementation fidelity acts as a potential moderator of the relationship between interventions and their intended outcomes. Unless implementation fidelity is assessed, in a circumstance of poor

outcome we cannot know whether the programme did not work or merely was not implemented properly, or both. Indeed, even in a circumstance of good outcome, we also cannot know whether the programme actually worked and was responsible for the positive outcome.

Researchers quickly found that measuring implementation fidelity was both complex and expensive. Not all educational interventions clearly specified what the teacher had to do and in what order. Indirect attempts which simply asked teachers whether they had implemented well were often found not to correlate with outcomes. Direct attempts which used observational methods to avoid teacher subjectivity were extremely expensive and still suffered from observer effects – what the teacher did when observed might not have been typical of what they did when not observed. Another issue was whether any professional development prior to the intervention was one-off, or whether it was several sessions with time in-between for reflection and discussion with colleagues, or whether it included ongoing coaching to shape teacher behaviour as the programme was being implemented. This latter raised an issue about just when implementation fidelity should be assessed – coupled of course with the issue of possible later implementation drift as the intervention continued over time.

The literature on implementation fidelity in book reading shows that direct and indirect methods are in general not strongly related to student outcomes. The present paper takes an alternative approach, investigating whether computerised methods examining student engagement with books are any more effective. It takes the indicators of implementation fidelity regarding student engagement in book reading determined empirically by Topping (2017), compares them with outcome indicators of student performance on reading achievement tests over a period, and asks which regions, local authorities and schools did best *on both* during the year in question. This latter is likely to be

important in practical terms, not only for regions and authorities which could be seen to be under-performing, but also to highlight schools which could be held up as examples of good practice.

The purpose of this paper was not to establish whether AR “works”. That issue has been addressed by other research (e.g. Siddiqui, Gorard, & See, 2015, which used a control group and found an effect size of +0.24). Any control group would need to be of pupils of the same reading ability, the same gender balance, in the same schools, but not contaminated by any contact with the programme. This paper is purely about implementation.

## **Previous Research**

### **Variation in Reading Outcomes**

Of course there are many studies on variation in reading outcomes by regions, states, local authorities, school districts and schools, and a few examples will be given here. Tennent, Stainthorp and Stuart (2008) sampled 11-year-old children attending all state-funded local authority schools in one London borough. Scripts of completed Standard Assessment Task reading papers were analysed to investigate the levels of inferential ability evident. Students were particularly weak in making inferences that require the application of background knowledge, and were consequently at risk when transferring to secondary education.

Similarly but on a larger scale, Sprietsma (2010) sought to estimate the effect of pupil's relative age within the first grade of primary school on mathematics and reading test scores at age 15 in 16 different countries. Relative age at the start of primary school had a significant effect on test scores in about one-half of the countries.

On an even larger scale, Dronkers and Kornda (2014) investigated gender differences in migrants from 62 countries using PISA data - 16,612 daughters and 16,804 sons of migrants in destination countries across Asia, Europe, Latin America, and Oceania. Female

migrant pupils had higher reading and math scores than comparable male migrant pupils, and these gender differences among migrant pupils were larger than among comparable native pupils.

Also on a large scale, Johansson, Myrberg and Rosén (2015) obtained data from pupils ( $n = 5271$ ) and teachers ( $n = 351$ ). Teacher competence was operationalized by multiple observed indicators. Two measures of achievement were used: PIRLS reading test results and teacher judgements of pupil performance. Teacher competence was positively related to both achievement measures. This study shows that measures which purport to be of student achievement may be indirectly also measuring teacher variables.

Returning to a smaller scale, Gorard, Siddiqui and See (2016) conducted an evaluation of Fresh Start (a synthetic phonics intervention) in ten secondary schools across England. Year 7 pupils (first year of secondary school) ( $n=433$ ) were identified as having insecure literacy attainment and individually randomised to a treatment group or a waiting-list control and assessed pre-post on the New Group Reading Test. The Effect Size was 0.24 and a sub-group of socio-economically disadvantaged students performed similarly.

These exemplar studies show that measuring student reading outcomes across schools, local authorities/school districts, regions/states and countries is common enough, and age in class and gender may be significant (although student reading outcomes may reflect a range of variables far beyond the student). What is less common is measuring variation in implementation fidelity in reading.

### **Variation in Implementation Fidelity in Reading**

McIntyre, et al. (2005) examined the implementation of ten early reading models through both structured observations and interviews. There was great variability in implementation fidelity. High implementers had much support, a practical, clear model, extensive professional development, or a combination of these. Similarly, Senesac and Burns

(2008) investigated the Help One Student to Succeed (HOSTS) Language Arts program. Participants were 51 elementary schools in urban areas and 1,354 students. Students in HOSTS programs that were implemented in a consistent manner achieved better reading outcomes.

In 2009, Zvoch investigated implementation fidelity in early childhood literacy programs delivered to economically disadvantaged students. He estimated protocol adherence at the onset as well as change in adherence over the intervention period (a relatively unusual step). Fidelity to protocol varied within and between treatment sites during the initial observation and over time. The background characteristics of teachers and contextual factors in the treatment environment were associated with fidelity outcomes. The results highlighted the challenge of achieving and maintaining fidelity to a treatment intervention that is delivered by multiple providers over multiple treatment sites.

An experimental evaluation of the Breakthrough to Literacy (BTL) program in Chicago was conducted by Unlu, et al. (2013). Forty-four schools were randomly assigned to either a treatment condition or a “business-as-usual” control condition (75 BTL and 58 control kindergarten teachers). In addition, all kindergarten students with data were included (1099 BTL and 787 control students). There was a positive association between the impact of BTL and its implementation fidelity.

In an intervention for struggling adolescent readers, Cantrell, Almasi, Carter and Rintamaa (2013) examined relationships among teachers' efficacy, implementation fidelity and students' reading progress. Findings indicated teacher efficacy was positively related to students' reading comprehension and overall reading achievement, while implementation fidelity was positively related to students' growth in vocabulary. This interesting study thus showed differential effects for standard and non-standard components of implementation.



Fogarty, et al. (2014) examined the effects of a multicomponent reading comprehension intervention (Comprehension Circuit Training - CCT) in sixth- to eighth-grade English language arts classes in three schools (14 teachers, 859 students), with a focus on implementation fidelity (specifically adherence, quality, dosage, program differentiation, and student responsiveness). Classes were randomly assigned to intervention or control conditions and all teachers taught in both conditions. Fidelity was significantly related to outcomes on a standardized comprehension measure ( $ES = 0.86$ ,  $p < 0.01$ ) and a narrative measure ( $ES = 0.52$ ,  $p < 0.01$ ). Results underscored the importance of measuring multiple dimensions of implementation fidelity.

An implementation fidelity study on Open Court Reading was conducted by Sullivan, Bell, Jones, Caverly and Vaden-Kiernan (2016). It distinguished between general fidelity (comparing intervention to control classes) and specific fidelity (variation within the intervention group), investigating two years of implementation in 49 elementary schools (kindergarten through 5th grade) in seven districts across the USA. None of the indicators for structural fidelity (dosage and adherence) were predictive of student reading outcomes. However, there was a relationship between teachers' process fidelity and students' achievement.

Thus, these studies explored variability in implementation fidelity across contexts of different sizes, from a handful of schools through school districts/local authorities and on to whole countries. Generally higher implementation fidelity was associated with better outcomes in achievement, but it is noteworthy that most studies only investigated teacher behaviour. High implementers tended to have strong support, a clear model and extensive professional development. Fidelity needed to be estimated during the whole of the intervention, since it might vary over time as well as between intervention sites. Finally,

teacher efficacy may be more important than implementation fidelity – but is also difficult to measure reliably. Let us turn now to implementation fidelity specifically in book reading.

### **Variation in Implementation Fidelity in Book Reading**

Unfortunately, extensive searches failed to find many significant studies of implementation fidelity specifically in book reading. There were almost none referring to variation among regions, states, local authorities, school districts, and schools. Studies linking implementation fidelity to reading achievement were equally scarce.

One relatively recent report (Renaissance Learning, 2012) analysed the data of 2,284,464 students in all 50 states in the USA who used Accelerated Reader (AR) (a programme measuring comprehension of books chosen for independent reading) and completed a STAR norm-referenced pre- and post- reading test. This was one of the first studies to investigate student response as an alternative to teacher behaviour. Implementation quality was estimated by Average Percent Correct (APC) on the quizzes. However, Time Spent Reading (ERT) and Challenge (ZPD) were estimated in a more convoluted manner. Of the mediating factors, APC appeared by far the most influential and ERT the least.

A recent UK study took a somewhat similar approach, but with different norms, using variables empirically derived as the most pertinent to UK data (see Topping, 2017). Reading outcome variables were highly correlated with each other at pre and post. Implementation variables also tended to be quite highly correlated with each other. However, outcome variables and implementation variables were not highly correlated with each other, implying that implementation fidelity contributed only modestly to pupil reading outcomes. Pupil Premium showed negative correlations with gains (as would be expected) but these were very small, suggesting that the effect of AR was not determined by SES. Primary children did far better on STAR reading test outcomes than secondary school children. They also did far better on implementation variables such as APC and AverageBookLevel-MidGP (a measure

of book difficulty or challenge) than secondary school pupils. Analysis of gender showed marked differences on both outcome measures and implementation fidelity measures, males being significantly worse. Students of higher reading ability in relation to age implemented AR at a higher level than students of lower ability. Despite this, they gained in reading at a lower level than students of lower ability (but only to a small degree).

The present study takes the Topping (2017) findings and further investigates whether there were marked differences between regions, local authorities and schools on these implementation and outcome measures.

### **Research Questions**

1. Which regions of the UK showed what levels of implementation fidelity of book reading and student achievement in reading?
2. Which local authorities in the regions of the UK showed what levels of implementation fidelity of book reading and student achievement in reading?
3. Which schools in the regions of the UK showed what levels of implementation fidelity of book reading and student achievement in reading?
4. Which schools in England showed the highest levels of implementation fidelity of book reading and student achievement in reading in relation to their level of socio-economic disadvantage as indicated by Pupil Premium Percentage?

### **Method**

#### **Sample**

The sample comprised all students in the UK for whom AR and STAR results were available for the academic year in question ( $n = 852,295$  in 3243 schools). This was 10.15% of the 8.4 million children in UK schools in 2015 (Department for Education, 2015). Schools

using AR only in primary numbered 1036 and schools using AR only in secondary numbered 1604. The number of schools using AR in both primary and secondary sectors (including middle schools and special schools) was 603, not allowing for the fact that all Scottish primary schools and some Northern Ireland schools have Year 7 students in primary school. Students in high schools outnumbered students in primary schools by three to one. However, data were not available on all variables for all students, since some schools were included which did not provide pre-post test scores for all classes/years of student in the school. Additionally, a decision was taken to disregard mid-year scores, which were very highly correlated with pre- and post-test scores. Consequently, some analyses were conducted on considerably fewer students. However, the number of students for each analysis was always large and is noted in the text.

## **Measures**

Accelerated Reader (AR) is a personalised practice and daily progress-monitoring system that helps teachers accurately and efficiently monitor pupil progress in quality (comprehension), quantity and difficulty of books read. First, a pupil reads a book at school, at home or elsewhere or in a variety of places, although all the cases here were subsequently assessed at school. The location(s) of reading were not measured for each book. In fact, independent reading time in school is limited, especially at secondary, so much of the reading probably did take place out of school. At the end of the book the pupil takes a computerised quiz of 5, 10, or 20 questions depending on the length of the book. Then, the pupil and teacher receive immediate computerised feedback with reports detailing books read, number of words read, book reading level and level of comprehension (Percent Correct on the quiz). AR was designed to make the job of managing book reading easier and more reliable whilst also motivating pupils to read more books for pleasure. The formative feedback helps

teachers shape subsequent reading instruction, guide individual pupils and motivate children to continue reading. Thus AR monitors both quantity and quality of individual book-reading.

STAR (Standardized Test for the Assessment of Reading) is a computerised standardized (norm-referenced) computer-adaptive item-banked reading test. Pupils respond to sentences with multiple-choice questions on a computer screen. The test is adaptive, i.e. it responds to the performance of each individual student. If the pupil succeeds, harder questions are given. If the pupil fails, easier questions are given. This greatly reduces testing time and student stress. The test is also item-banked, i.e. it has multiple items at the same level. Consequently students cannot copy from each other as no-one is doing the same test. This also enables the test to be taken frequently without practice effects. On completion feedback is available immediately to the teacher and/or pupil. STAR Reading has test-retest reliability of 0.92, split-half reliability of 0.91 and generic reliability of 0.97 in the US. Generic reliability in the UK is 0.94. In term of validity, STAR Reading correlates at 0.96 with the Degrees of Reading Power test. Predictive and concurrent validity with a great number of other reading tests are reported (Renaissance Learning, 2013, 2014).

### **Definitions**

Scaled Score (SS) ranges from 0 to 1400 and spans years 1–13. It is calculated based on the difficulty of the questions and the number of correct responses. Scaled Scores are useful for comparing student performance over time and across years.

Student Growth Percentile (SGP) (Betebenner, 2011) is taken from the SS scores on two or more tests within 18 months to give an indication of the student's growth trajectory. The SGP is a norm-referenced percentile-based index derived using quantile regression techniques to establish curvilinear functional relationships between the cohort's prior scores and the cohort's current scores. It ranges from 1 to 99 and indicates how exemplary a student's growth from one test window to another is relative to students in the same grade

with a similar achievement history across the US. SGPs have a national median of 50. An SGP of 10, for example, would indicate growth of that student which exceeded 10 percent of their academic peers' growth and was less than 90 percent of their academic peers, i.e. relatively low. Conversely, an SGP of 90 would indicate growth exceeding 90 percent of their academic peers. SGP percentiles are robust to outliers and uncorrelated with prior achievement. Because SGP is a mathematical manipulation, normal issues of reliability and validity do not apply. Instead, issues of accuracy and precision apply. Shang, VanIwaarden and Betebenner (2015) found that SGP tends to be overestimated among students with higher prior achievement and underestimated among those with lower prior achievement. The simulation-extrapolation method known as SIMEX was used to correct these anomalies. Approximately 10% of classes at the extremes were affected. Wright (2010) noted that SGPs correlated highly with value-added models but both under-estimated high-poverty classrooms, with SGP under-estimating least.

Average Percent Correct (APC) is the percentage of correctness of the student's answers to the quiz questions, aggregated over all books the student has read.

Grade Placement (GP) is a numeric representation of a student's grade level, based on the specific month in which a student takes a STAR Reading test. STAR Reading considers the standard school year to run from September through June and assigns increment values of 0.0 through 0.9 to these months. The software automatically assigns grade placements using a student's grade level and the month in which a STAR Reading test was taken. GP thus expresses the National Curriculum Year (to one decimal place) in the English context. Scottish grades have been converted to the English system.

AverageBookLevel-MidGP (ABL-MidGP) was a derived variable intended to indicate the degree of challenge in the books each student was reading. The Average Book Level per student was available in the dataset, being the average level of difficulty of the

books each student read during the year, expressed as a grade. The Average Book Level was determined by the ATOS formula, which applied to an analysis of the whole book, not merely selected passages. From this was subtracted the chronological age (or more precisely, the Grade Placement in years and months) of each student. As the Average Book Level covered the whole year, the Grade Placement at the mid-point of the year was compared to it. In the early years, ABL-MidGP was positive as Average Book Level was higher than Grade Placement, but in later years the opposite was true.

Pupil Premium is additional funding for publicly funded schools only in England, intended to raise the attainment of disadvantaged pupils and close the gap between them and their peers. It is allocated to those who have been eligible for Free School Meals at any point in the last six years, children who are looked after by the local authority, and children whose parents are currently serving in the armed forces. The percentage of pupils in the school for whom the premium is received is the variable.

### **Data analysis**

Regions were compared using Student's t-test. It could be argued that the sample was not randomly selected and that therefore this test was not applicable. However, the size of sample was so large that it was thought reasonable to apply this analysis. The comparisons by local authority and school were done on the basis of descriptive statistics, with reference to the number of cases providing data.

## **Results**

### **Differences by region of the UK**

As Table 1 shows, Northern Ireland achieved the highest achievement gains in both SS and SGP, significantly ahead of England. Northern Ireland was also significantly ahead of Scotland and Wales, by even larger margins. England was significantly ahead of Scotland

and Wales. Scotland was ahead of Wales but the difference is not statistically significant, possibly owing to the smaller numbers in the two samples.

#### INSERT TABLE 1 ABOUT HERE

Now we can look at whether these differences in gains were matched by differences in quality of implementation, as indicated by the two main variables of importance: APC and ABL-MidGP (see Table 2). Northern Ireland scored significantly higher on APC and ABL-MidGP than England, so the difference in outcomes between Northern Ireland and England appears to be attributable to differences in implementation. However, the difference in outcomes between Scotland and England does not appear to be attributable to differences in implementation. Scotland scored significantly higher on APC and ABL-MidGP than England, but still had worse outcomes. England and Wales showed insignificantly different levels of APC, but ABL-MidGP was higher for England. The outcomes in England were higher than in Wales. Northern Ireland implemented significantly better than Scotland and had significantly better outcome scores. Northern Ireland implemented much better than Wales and had significantly better outcome scores than Wales. Scotland implemented significantly better than Wales on both variables, but Scotland's outcome scores were not significantly different from those in Wales.

#### INSERT TABLE 2 ABOUT HERE

The general pattern is of implementation quality being highly related to outcome scores. The exception to this is Scotland, where relatively high implementation scores do not correspond with high outcome scores. It seems that at the regional level of analysis, a stronger relationship between implementation and outcomes is present than when all data are considered together using another form of statistical analysis (as in Topping, 2017). Of course this does not tell us what is happening on the ground, but discovering that would be a very large task.



### **Differences by local authority**

The key outcome indicators (SSGain and SGP) and implementation indicators (APC and ABL-MidGP) were available for many of the local authorities in the UK (Table 3). These are divided into results for Wales, Scotland, Northern Ireland and England separately, with an indication of the mean of each variable for each region. However, in some cases (in Wales and Scotland) these data were available for a very small subset of the pupils in the database and so these authorities must be disregarded (although they appear below for purposes of completeness). Mean Pupil Premium has been added for authorities in England, so that it becomes possible to see the extent to which socio-economic disadvantage is a factor in these scores.

#### INSERT TABLE 3 ABOUT HERE

Considering the means per region, again Scotland performed worst on outcomes, although it had the second highest APC and the highest ABL-MidGP. Northern Ireland performed exceptionally well on both outcomes and implementation. On this analysis, Wales and England did equally well on both outcomes and implementation.

In Wales, Clwyd and Mid Glamorgan both did relatively well in outcomes and implementation, but pupils in Mid Glamorgan were more likely to read books of difficulty below their chronological age. West Glamorgan had a lower SSGain but higher SGP than either, and a considerably higher APC and ABL-midGP, similar to Clywd.

In Scotland, West Lothian was strikingly the best authority with a substantial proportion of pupils in the sample, with the best outcomes and highest level of difficulty of books. APC was not however quite as high as elsewhere. Kirkcudbright also did well, with only slightly lower outcomes, APC and book difficulty. Dumfriesshire, Wigtownshire and Caithness followed. Largely rural authorities were in preponderance. Renfrewshire came next,

an urban/rural mixture achieving good results. The best authorities in Scotland had considerably higher SSGain and ABL-midGP than the best in Wales.

In Northern Ireland there were a large number of schools with missing data. County Tyrone did best, with high SSGain and low ABL-midGP, comparable to West Lothian in Scotland. County Armagh came next but with a higher ABL-midGP. Beyond these two largely rural authorities, County Derry had very high outcome scores and a very low ABL-midGP, but with a smaller proportion of children yielding data.

In England, Cambridgeshire and Suffolk performed extremely well, with SSGains over 90 and ABL-midGP small. After them, Warwickshire also had high outcomes and low ABL-midGP. Bedfordshire, Berkshire, Buckinghamshire, Cornwall, County Durham, Cumbria, Devon, Dorset, East Sussex, Herefordshire, Hertfordshire, Kent, Shropshire, Somerset and South Gloucestershire showed high SSGain, but APC and ABL-midGP were often relatively low. Many of these authorities were largely rural. London did particularly well in the light of its high Pupil Premium score, with high gains and moderate ABL-midGP (but low APC). Of the others, only County Durham and London had a Pupil Premium Percentage substantially above the national average. Humberside (not to be confused with North Humberside or South Humberside), with a very high Pupil Premium, had only a small proportion of cases yielding data, but appeared to show very high SSGain and high APC, albeit with a low ABL-midGP. The variation in ABL-midGP was quite striking, and seemed not to relate to urban/rural status or any other obvious variable.

### **Analysis by School**

The total number of schools yielding data for all four regions in the school analysis was 1989, excluding schools which had mid-year but no post-test data, special schools and schools whose data were unusable. It was interesting to see in each of the regions of the UK, in primary and secondary schools separately, which schools seemed to be implementing AR

well and achieving high outcome gains. Obviously, such schools could be held up as centres of good practice for other schools to visit. We initially decided to identify the best five schools at both primary and secondary level in each of the four regions of the UK. All schools were identified only by their numerical school identity number during this process, so it was carried out blind to the identity of the schools.

Selection criteria were established. We eliminated schools with a small number of students (below 20), as we felt success with a more substantial number of pupils was important. Secondly, we eliminated schools who had many pupils for whom the data was incomplete (below 66%), as we did not wish to make judgements based on what might be a biased sample of pupils from the school. Schools were coded as primary, secondary or special. Special schools and the few schools which could not be identified were then excluded. For this analysis, comparisons were made only between primary and secondary, so pupils in middle schools were categorised primary or secondary according to the Year they were in. Pupils in schools catering for the whole age range (primary plus secondary) were similarly categorised as primary or secondary according to their Year group. Most schools were state schools but some were private schools.

We gave priority to SGP and APC (over SSGain and ABL-midGP) as the most important indicators when making choices at the margin. When conducting analyses for schools in England we were able to add the Pupil Premium Percentage, to give an indication of the extent to which socio-economic disadvantage might be playing a part in the school's results, and the extent to which the school had overcome this. We will give the **school** results for each region in turn, starting with primary and then adding secondary.

Wales unfortunately only had five schools yielding data at primary level. Of these, one did not meet the criterion for the number of participating pupils and one did not meet the

criterion for the proportion of pupils included. However, we give the ranked list for the remaining three schools in Table 4.

INSERT TABLE 4 ABOUT HERE

At secondary level, Wales had 16 secondary schools yielding data. The top five are given in Table 5.

INSERT TABLE 5 ABOUT HERE

Scotland had 101 primary schools yielding data. The top five primary schools in Scotland are as described in Table 6.

INSERT TABLE 6 ABOUT HERE

Scotland had 16 secondary schools yielding data. The top five secondary schools in Scotland are described in Table 7. Some unusual aspects are evident in this table. School 1894975 had high SGP gain but low APC, which is why it is in second rather than first place. School 2197856 had a very low APC but the level of difficulty of books was very high. School 2006479 had a low APC but the level of difficulty of books was very low. However, this school has a Free School Meals percentage of 40%, compared to 27% for the rest of Glasgow and 15% for the whole of Scotland. Its appearance as a top 5 school is highly creditable.

INSERT TABLE 7 ABOUT HERE

Northern Ireland had 159 primary schools yielding data. The top five primary schools in Northern Ireland are described in Table 8. The Northern Ireland data were characterised by the large number of schools who had high numbers and percentages of children participating. They were also characterised by many high SGP scores. Consequently, identifying the top 5 schools is somewhat invidious. School 2235026 combined this with reading very difficult books.

INSERT TABLE 8 ABOUT HERE

Northern Ireland had 36 secondary schools yielding data. The top five secondary schools in Northern Ireland are as described in Table 9. School 2259586 had a low APC with hard books while school 2252017 had a high APC with easier books (although no easier than the other schools who had a lower APC). It appeared that the prominence of Northern Ireland in the previous region comparison was largely due to their highly successful primary schools.

INSERT TABLE 9 ABOUT HERE

England had 457 primary schools yielding data. This was a large number and in the interests of representativeness a decision was taken to list all schools with SGP greater than or equal to 60. This yielded 40 primary schools (Table 10). APC was very variable in these schools. Only 12 schools had a high APC. Some schools with very high SGP and SSGain had very low APC. ABL-MidGP was also variable, although four schools showed a positive figure. The highest was 1.79, suggesting pupils at this school tended to read easy books. The appearance of schools 1893902 and 1892270 high up the ranked list despite their large Pupil Premiums is highly creditable. Schools 1893701, 1950401, 2240654, 2235716 and 2235887 also do well in this regard. The mean Pupil Premium for all English primary schools was 28 (range 0-74).

INSERT TABLE 10 ABOUT HERE

England had 1199 secondary schools yielding data. Again, the schools with a SGP of 60 or more were listed. This yielded the top 19 secondary schools in England (Table 11). Four of the 19 were grammar schools. Four were schools only for boys, which may have constituted a disadvantage as boys are generally less able on reading tests. Seven of these 19 schools had an APC at or above the recommended level, while 12 did not. ABL-MidGP generally hovered around -2.5, but some schools were below this and in two cases considerably below. Pupil Premiums in these schools were very various, ranging from 0 to 50. The average Pupil Premium for secondary schools in England was 29.13 (range 0-100).

Three schools stood out as having Pupil Premiums well above this average - their performance was highly creditable. This is particularly true as SGP tends to under-estimate the scores of disadvantaged schools. It seems surprising that school 3385205 has such a high SGP and SSGain. It also seems surprising that schools 2142013, 2248873 and 2237433 have such a large SGP with such a small SSGain.

INSERT TABLE 11 ABOUT HERE

## **Discussion**

### **Summary**

Differences between Regions of the UK were explored. Northern Ireland achieved the highest outcome gains in both SS and SGP, significantly ahead of England. Northern Ireland was also significantly ahead of Scotland and Wales, by even larger margins. England was significantly ahead of Scotland and Wales. Scotland was ahead of Wales but the difference was not statistically significant, possibly owing to the smaller numbers in the two samples. The general pattern was of implementation quality being highly related to outcome scores. The exception to this was Scotland, where relatively high implementation scores did not correspond with high outcome scores.

Differences between local authorities were then explored. These were divided into the results for Wales, Scotland, Northern Ireland and England separately, with an indication of the mean of each variable for each region to enable comparison. Mean Pupil Premium was added for authorities in England, so that it became possible to see the extent to which socio-economic disadvantage was a factor. The highest performing authorities were largely rural. Exceptions were Renfrewshire, County Derry, County Durham, Humberside and London. ABL-midGP seemed to vary considerably between authorities but not in any predictable way.

The results of individual schools were then explored, to establish which schools in each of the four regions (England, Wales, Scotland, Northern Ireland) had the best results in terms of outcomes and implementation. Such schools could be held up as centres of good practice for other schools to visit. We initially decided to identify the best five schools at both primary and secondary level in each of the four regions of the UK. We eliminated schools with a small number of students (below 20) and those who had many pupils for whom the data was incomplete (below 66%). Results were divided in primary and secondary (pupils in middle schools and all-through schools were coded primary or secondary by their Year). Pupil Premium Percentage was added for England. The top five schools were identified in

Scotland, Wales and Northern Ireland. However, schools in England were so much more numerous it was decided to include all schools with SGP above the 60<sup>th</sup> percentile.

In Northern Ireland, all five Primary schools had a high APC with SGPs ranging from 62-73, while 2/5 Secondary schools had a high APC with SGPs ranging from 50-57. In England, 12/40 Primary schools had high APCs while SGPs ranged from 60-75, while 7/19 Secondary schools had high APCs with SGP ranging from 60-85. In Scotland, 3/5 Primary schools had a high APC with SGPs ranging from 53-59, while 1/5 Secondary schools had a high APC with SGPs ranging from 43-53. In Wales, Primary SGPs ranged from 52-63, Secondary from 41-55. Some schools appeared high up these ranked lists despite a high Pupil Premium.

Overall, primary pupils read more books but these books were relatively hard in relation to the chronological age of the pupils, while secondary pupils read fewer books (partly because these books are longer) but these books were relatively easy in relation to the chronological age of the pupils.

### **Relevance to previous literature**

The present study was one of few utilising student responses as an implementation variable, most of the previous studies having focused on teacher behaviour. It found that although student responses were only as good as teacher behaviour when it came to relationships with outcomes (Topping, 2017), when data were inspected at a more micro level (regions, local authorities, individual schools) a stronger relationship between implementation and outcomes emerged.

We have also seen from the previous literature that few studies of reading outcomes and implementation fidelity reported implementation fidelity over a longer period, such as was done here. Even fewer reported implementation fidelity indices available as a matter of course without additional effort during the implementation of an intervention such as AR which occurs during many consecutive school years. Neither direct nor indirect methods of assessing implementation fidelity included any computer-based measures. Turning to previous reports on AR and STAR, we find that APC was noted as the strongest variable, which is also what was found here.

### **Limitations**

The present study had a number of limitations, as well as a number of advantages. The principal advantage was the large sample size, although much larger samples are available in the US. Sample size was reduced by some schools entering pupils for testing but then not providing results on them all. It was also reduced by the discarding of mid-session data, which excluded those schools which only provided pre- and mid- data. Nonetheless, the sample size remained large, certainly much larger than in most studies of reading.

Additionally, we have noted that SGP tends to under-estimate schools in socio-economically disadvantaged areas and over-estimate schools in advantaged areas. This suggests that when interpreting the tables, readers should judge flexibly in the top and bottom



quintiles. Given this, it is surprising that so many schools with high Pupil Premium did well in our ranked lists of top scoring schools.

The analysis by local authority had the difficulty that the balance between primary and secondary schools is different in different authorities. Some authorities may have a larger number of primary schools represented, which may mean that they are more likely to have better results than other authorities. Additionally, where a large number of schools from the authority are included, they are more likely to be representative of the authority than where a small number of schools are included. In some cases the data were available for a very small subset of the pupils in the database, and so results for these authorities must be disregarded. It may also be that some of the very high SSGain and SGP scores reflect a situation where schools have only just implemented AR and are experiencing a surge in their outcomes, which will slow to a more sustainable level in ensuing years.

There are also issues with the analysis by school, since some schools may target AR only where they think it will be most effective, while other schools may offer it to the whole school. The schools that do the former may tend to have better results than those that do the latter.

### **Implications for practice, policy and future research**

**Practice.** The schools listed in the foregoing top performing lists are doing well. Beyond these schools, there are clearly many schools where quality of implementation falls short. Topping (2017) pointed out that teachers should strive to maximise implementation of what appear to be the major determinants of higher outcomes from empirical analysis – *Accuracy* on quizzes (i.e. Average Percent Correct, APC) and *Challenge* (Average Book Level – MidGP). Of course, teachers are working indirectly with students who generate the data, so much of their work will involve explaining this to students and subsequently coaching them in relation to their own performance indicators. When teachers evaluate the success of AR in

their schools, they should carefully consider the current evidence on these two key indicators of implementation fidelity as well as the level of student outcomes, and strive to increase them.

**Policy.** At local authority and regional level, policymakers should inspect their local authority and regional level of performance and consider how it might be improved. It may well be that investment in a process of professional development which continues over time and is supported by coaching and further inspection of results will bear fruit.

**Future research.** If studies are repeated, it would be useful to investigate the two key empirically determined implementation fidelity variables (APC and ABL-midGP). In the UK, further work could seek to incorporate the mid-year STAR data. It may well be that the regions, local authorities and schools will show some change over time (as a result of spontaneous as well as planned changes), and thus there is a case for repetition of this research every year. More generally, however, it seems that computer assisted assessment of implementation fidelity is an interesting alternative to other forms of measuring implementation integrity which focus exclusively on teacher behaviour, and its use in other areas of the curriculum should be explored.

### Conclusion

In the regional analysis, the general pattern is of implementation quality being highly related to outcome scores. The exception to this is Scotland, where relatively high implementation scores do not correspond with high outcome scores. Northern Ireland did better than England (mainly because of high-performing primary schools), while England did better than Scotland and Wales.

In the local authority analysis, the highest performing authorities were largely rural. Exceptions were Renfrewshire, County Derry, County Durham, Humberside and London. ABL seemed to vary considerably between authorities but not in any predictable way.

In the school analysis, insufficient primary schools were available in Wales. In secondary schools, outcomes were very variable and ABL poor. In Scottish primary schools, the best schools had high gains and often also high implementation. There are clearly examples of excellent performance in Scotland which other primary schools would do well to follow. In secondary gains were lower and ABL-midGP tended to be poorer, although APC was good in some cases. In Northern Ireland, outcomes and APC were very high, with ABL-midGP not so strikingly high as in Scotland. The top secondary schools had lower and much more varied outcomes, with wide variation in APC and considerably less good ABL-midGP. England yielded an extensive list of high performing primary schools, with very high outcomes and generally good ABL-midGP but very variable APC. While some schools had high APC, others were lower.

In England most of these primary schools had Pupil Premium below the average, but some high up the list did not, and must be commended on their excellent performance. This was also true in secondary, with three schools with high Pupil Premium performing excellently out of 19. SSGains were also high in the secondary schools, although SGP gave a more modest picture. APC was high in five cases (out of 19), but the worst case was in the 60s. ABL-midGP was also variable, ranging from 2 to 4.6.

A number of private schools appeared in these lists (except in Northern Ireland). It is not clear whether we might expect private schools to do well compared to state schools (as they tend to be more advantaged), or whether we should be surprised to see them in these lists, since the AR appears to have made a major difference to what would have been assumed to be an already advantaged school.

With reference to the research questions, then, this study has demonstrated which regions of the UK showed what levels of implementation fidelity of book reading and student achievement in reading (Northern Ireland best on both, England next best on both, Scotland high on implementation but low on achievement, and Wales lowest on both). It also showed which local authorities reliably showed what levels of implementation fidelity of book reading and student achievement in reading, which schools in the regions of the UK reliably showed what levels of implementation fidelity of book reading and student achievement in reading, and which schools in England showed the highest levels of implementation fidelity of book reading and student achievement in reading in relation to their level of socio-economic disadvantage as indicated by Pupil Premium Percentage.

Computerised methods are an interesting alternative for investigating the relationship between implementation fidelity and student outcome in book reading. This method of measuring implementation fidelity focusing on student performance is in sharp contrast to past studies of implementation fidelity of reading, which have tended to focus on teacher behaviour rather than student performance. The relationship between implementation quality as measured by AR and reading outcome as measured by STAR is not strong when all data are assessed together (as in Topping, 2017), which puts this measure of implementation fidelity at about the same level as other measures in the literature which predominantly look at teacher behaviour. However, when data are examined by region, local authority and school a stronger relationship between implementation and outcomes appears which has important implications for practice.

Certainly, some way of measuring both student engagement and teacher behaviour would be desirable. The present difficulty is that measures of teacher behaviour would probably need to be self-report, involving the low reliability associated with such measures, as computerised measures of relevant teacher behaviour are not yet available. While AR

could possibly be configured to measure the number of times and length of time teachers accessed feedback, whether this was thought about deeply or acted upon is another issue.

Future research needs to investigate reliable and valid measures of teacher behaviour which can be coupled with existing measures of student engagement.

## References

- Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories*. Dover, New Hampshire: The National Center for the Improvement of Educational Assessment. Retrieved from [http://www.nj.gov/education/njsmart/performance/SGP\\_Technical\\_Overview.pdf](http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf) [14 June 2016].
- Cantrell, S. C., Almasi, J. F., Carter, J. C., & Rintamaa, M. (2013). Reading intervention in middle and high schools: implementation fidelity, teacher efficacy, and student achievement. *Reading Psychology*, 34, 26-58.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 72, 40. Retrieved from <https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-2-40> [14 June 2016]. DOI: 10.1186/1748-5908-2-40.
- Department for Education (2015). *Schools, pupils and their characteristics: January 2015*. London: DfE.
- Dronkers, J., & Kornder, N. (2014). Do migrant girls perform better than migrant boys? Deviant gender differences between the reading scores of 15-year-old children of migrants compared to native pupils. *Educational Research and Evaluation*, 20(1), 44-66.
- Durlak, J. P., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327-350.
- Fogarty, M., Oslund, E., Simmons, D., Davis, J., Simmons, L., Anderson, L., Clemens, N., & Roberts, G. (2014). Examining the effectiveness of a multicomponent reading

comprehension intervention in middle schools: A focus on treatment fidelity.

*Educational Psychology Review*, 26, 425-449.

Gorard, S., Siddiqui, N., & See, B. H. (2016). An evaluation of Fresh Start as a catch-up

intervention: A trial conducted by teachers. *Educational Studies*, 42(1), 98-113.

Johansson, S., Myrberg, E., & Rosén, M. (2015). Formal teacher competence and its effect on

pupil reading achievement. *Scandinavian Journal of Educational Research*, 59, 564-582.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity:

An essential—but often forgotten—component of school-based interventions.

*Preventing School Failure*, 48(3), 36–43.

McIntyre, E., Powell, R., Coots, K. B., Jones, D., Powers, S., Deeters, F., & Petrosko, J.

(2005). Reading instruction in the NCLB Era: Teachers' implementation fidelity of

early reading models. *Journal of Educational Research & Policy Studies*, 5(2), 66-102.

Renaissance Learning (2012). *Guided independent reading*. Wisconsin Rapids, WI:

Renaissance Learning. Retrieved from

<http://doc.renlearn.com/KMNet/R005577721AC3667.pdf> [16 June 2016].

Renaissance Learning (2013). *STAR Reading Technical Manual (UK)*. London: Renaissance

Learning.

Renaissance Learning (2014). *The research foundation for STAR assessments*. Wisconsin

Rapids, WI: Renaissance Learning. Retrieved from

<http://doc.renlearn.com/KMNet/R001480701GCFBB9.pdf> [17 June 2016].

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research:

Multidisciplinary perspectives on the conceptualization, measurement, and

enhancement of treatment integrity. *School Psychology Review*, 38, 460–475.

- Senesac, B. V., & Burns, M. K. (2008). Theoretical soundness, proven effectiveness, and implementation fidelity of the HOSTS language arts program among children identified as at-risk in urban elementary schools. *Journal of Instructional Psychology*, 35, 212-221.
- Shang, Y., VanIwaarden, A., & Betebenner, D. W. (2015). Covariate measurement error correction for Student Growth Percentiles using the SIMEX method. *Educational Measurement: Issues and Practice*, 34(1), 4-14.
- Siddiqui, N., Gorard, S., & See, B. H. (2015). Accelerated Reader as a literacy catch-up intervention during primary to secondary school transition phase. *Educational Review*, 68(2), 139-154. <http://dx.doi.org/10.1080/00131911.2015.1067883>
- Sprietsma, M. (2010). Effect of relative age in the first grade of primary school on long-term scholastic results: International comparative evidence using PISA 2003. *Education Economics*, 18(1), 1-32.
- Sullivan, K., Bell, N., Jones, D. H., Caverly, S., & Vaden-Kiernan, M. (2016). Implementation work at scale: An examination of the fidelity of implementation study of the scale-up effectiveness trial of Open Court Reading. *Society for Research on Educational Effectiveness*. ERIC Number: ED567238. Retrieved from <http://files.eric.ed.gov/fulltext/ED567238.pdf> [22 August 2016].
- Tennent, W., Stainthorp, R., & Stuart, M. (2008). Assessing reading at Key Stage 2: SATs as measures of children's inferential abilities. *British Educational Research Journal*, 34, 431-446.
- Topping, K. J. (2017). Theoretical and empirical indicators of implementation integrity in book reading. *Paper submitted for publication*.
- Unlu, F., Bozzi, L., Layzer, C., Smith, A., Price, C., & Hurtig, R.. (2013). Linking implementation fidelity to impacts in an RCT. *Society for Research on Educational*



*Effectiveness*. ERIC Number: ED563378. Retrieved from

<http://files.eric.ed.gov/fulltext/ED563378.pdf> [22 August 2016].

Wright, S. P. (2010). *An investigation of two nonparametric regression models for value-*

*added assessment in education*. Cary, NC: SAS Institute Inc. Retrieved from

[https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-](https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-Report-Cards/Value-Added-Technical-Reports-1/An-Investigation-of-Two-Nonparametric-Regression-Models-for-Value-Added-Assessment-in-Education-S-Paul-Wright-1.pdf.aspx)

[Report-Cards/Value-Added-Technical-Reports-1/An-Investigation-of-Two-](https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-Report-Cards/Value-Added-Technical-Reports-1/An-Investigation-of-Two-Nonparametric-Regression-Models-for-Value-Added-Assessment-in-Education-S-Paul-Wright-1.pdf.aspx)

[Nonparametric-Regression-Models-for-Value-Added-Assessment-in-Education-S-](https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-Report-Cards/Value-Added-Technical-Reports-1/An-Investigation-of-Two-Nonparametric-Regression-Models-for-Value-Added-Assessment-in-Education-S-Paul-Wright-1.pdf.aspx)

[Paul-Wright-1.pdf.aspx](https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-Report-Cards/Value-Added-Technical-Reports-1/An-Investigation-of-Two-Nonparametric-Regression-Models-for-Value-Added-Assessment-in-Education-S-Paul-Wright-1.pdf.aspx) [16 June 2016].

Zvoch, K. (2009). Treatment fidelity in multisite evaluation: A multilevel longitudinal

examination of provider adherence status and change. *American Journal of*

*Evaluation*, 30, 44–61.

Table 1: Mean differences in SSGain and SGP between regions of the UK and statistical significance

	SSGain					SGP				
	mean	sd	n	<i>t</i>	p	mean	sd	n	<i>t</i>	p
England	76.56	144.863	419747	-7.539	<.001	47.84	29.278	412607	-11.881	<.001
N. Ireland	84.54	126.210	19377			50.39	28.398	19349		
England	76.56	144.863	419747	4.842	<.001	47.84	29.278	412607	8.259	<.001
Scotland	68.82	129.993	8352			45.16	29.615	8348		
England	76.56	144.863	419747	5.201	<.001	47.84	29.278	412607	4.484	<.001
Wales	64.56	142.854	3982			45.74	28.973	3967		
N Ireland	84.54	126.210	19377	9.451	<.001	50.39	28.398	19349	13.880	<.001
Scotland	68.82	128.993	8352			45.16	29.615	8348		
N Ireland	84.54	126.210	19377	8.886	<.001	50.39	28.398	19349	9.361	<.001
Wales	64.56	142.854	3982			45.74	28.973	3967		
Scotland	68.82	128.993	8352	1.655	.098	45.16	29.615	8348	-1.022	.307
Wales	64.56	142.854	3982			45.74	28.973	3967		

Table 2: Mean differences in APC and ABL-MidGP between regions of the UK and statistical significance

	APC					ABL-MidGP				
	mean	sd	n	<i>t</i>	p	mean	sd	n	<i>t</i>	p
England	73.898	17.968	619777	-44.656	<.001	-2.881	1.745	480313	-45.940	<.001
N. Ireland	78.119	15.961	37879			-2.373	1.559	26033		
England	73.898	17.968	619777	24.074	<.001	-2.881	1.745	480313	-77.807	<.001
Scotland	77.204	17.263	17575			-1.620	1.765	11881		
England	73.898	17.968	619777	1.300	.194	-2.881	1.745	480313	30.432	<.001
Wales	73.627	18.753	7496			-3.605	1.458	5415		
N Ireland	78.119	15.961	37879	6.123	<.001	-2.373	1.559	26033	-41.853	<.001
Scotland	77.204	17.263	17575			-1.620	1.765	11881		
N Ireland	78.119	15.961	37879	21.598	<.001	-2.373	1.559	26033	53.479	<.001
Wales	73.627	18.753	7496			-3.605	1.458	5415		
Scotland	77.204	17.263	17575	14.631	<.001	-1.620	1.765	11881	72.274	<.001
Wales	73.627	18.753	7496			-3.605	1.458	5415		

Table 3: Outcomes and implementation by local authority per region

## Wales

Authority	Cases with data	% with data	SS Gain mean	SGP mean	APC mean	ABL mean
Clwyd	958	52.6%	68.54	43.89	72.378	-3.478
Dyfed	315	36.7%	47.76	42.26	79.076	-2.888
Gwent	83	10.0%	115.63	61.70	81.162	-2.228
Gwynedd*	23	13.5%	142.39	62.61	70.241	-2.401
Mid Glamorgan	1070	41.4%	69.86	46.47	73.790	-4.033
South Glamorgan	664	23.3%	52.55	43.42	77.750	-3.971
West Glamorgan	854	48.4%	62.65	48.00	80.727	-3.417
mean			79.91	49.76	76.45	-3.202

\*data available for a very small subset of pupils

## Scotland

Authority	Cases with data	% with data	SS Gain mean	SGP mean	APC mean	ABL mean
Aberdeenshire	1707	43.3%	57.38	43.35	78.250	-2.276
Argyll	183	32.3%	58.64	44.89	84.561	-1.173
Ayrshire	115	7.8%	55.47	37.63	73.127	-1.289
Banffshire	197	29.7%	79.95	46.13	75.086	-1.451
Caithness	292	79.8%	69.54	44.27	82.984	-1.467
Dumfriesshire	1026	50.7%	76.19	46.17	82.747	-1.636
East Lothian	264	37.0%	102.19	53.95	81.044	-1.359
Fife	513	34.9%	58.42	41.01	79.025	-1.294
Inverness-Shire*	72	94.7%			88.312	.051
Isle of Mull*	15	93.8%	138.53	62.20	89.070	-.853
Kincardineshire*	49	89.1%	99.73	47.94	85.759	-.498
Kirkcudbright	417	72.9%	83.54	48.46	77.518	-1.258
Lanarkshire	502	27.6%	75.85	46.88	76.007	-2.551
Midlothian	294	21.4%	58.40	41.23	82.101	-1.931
Morayshire	250	40.8%	69.29	41.23	77.642	-1.602
Perthshire	365	44.5%	79.15	47.19	83.952	-1.455
Renfrewshire	551	49.2%	66.57	43.49	82.275	-1.511
Ross-shire*	13	8.1%	-4.62	25.38	73.590	-3.278
Selkirkshire*	15	11.5%	116.20	49.27	81.095	-1.044
Shetland Isles*	58	45.7%	69.14	46.97	76.781	-1.807
Stirlingshire	659	48.7%	47.47	44.71	73.175	-.655
West Lothian	353	54.7%	95.90	52.80	79.501	-1.162
Wigtownshire	480	62.2%	75.77	45.94	78.633	-1.537
mean+			74.03	44.28	75.984	-1.504

\*data available for a very small subset of pupils

+Inverness-Shire omitted

#### Northern Ireland

Authority	Cases with data	% with data	SS Gain mean	SGP mean	APC mean	ABL mean
County Antrim	7712	41.6%	80.63	49.35	78.532	-2.545
County Armagh	1353	62.2%	89.63	52.12	79.742	-2.213
County Derry	194	41.1%	144.55	59.18	80.536	-3.294
County Down	2564	49.0%	77.51	50.67	77.070	-3.331
County Ferman	628	34.9%	85.40	48.89	81.286	-1.973
County London	3874	39.6%	85.67	50.48	81.868	-2.125
County Tyrone	3024	54.9%	92.73	51.69	82.071	-1.563
mean			93.73	51.77	80.158	-2.435

#### England

Authority	Cases with data	% with data	SS Gain mean	SGP mean	APC mean	ABL mean	Mean Pupil Premium
Avon	3356	56.9%	70.78	46.94	77.686	-2.874	22.96
Bedfordshire	9091	58.3%	84.46	49.67	75.693	-2.573	27.87
Berkshire	5172	59.2%	88.32	50.92	74.775	-2.922	21.53
Buckinghamshire	6513	57.6%	84.56	50.60	75.599	-2.533	17.40
Cambridgeshire	6991	54.4%	90.75	48.92	79.489	-1.375	25.31
Cheshire	10593	55.1%	69.60	47.33	75.827	-3.427	24.86
Cleveland	5925	55.0%	68.03	45.75	76.954	-3.501	37.47
Cornwall	5492	58.2%	88.48	48.79	78.460	-2.258	26.21
County Durham	7992	59.6%	81.31	47.02	78.977	-2.858	34.67
Cumbria	6909	65.1%	85.41	49.40	78.226	-2.428	20.79
Derbyshire	3488	61.9%	72.33	45.40	76.991	-2.752	28.27
Devon	16361	56.3%	86.39	49.49	78.476	-2.175	22.56
Dorset	1888	41.7%	81.84	50.25	77.636	-2.688	16.29
East Sussex	4161	41.4%	82.99	48.84	76.579	-2.438	27.59
East Yorkshire	450	86.2%	40.74	39.20	77.698	-3.466	19.25
Essex	27312	45.5%	72.98	46.73	75.982	-3.014	29.78
Gloucestershire	4275	50.0%	79.06	48.40	75.457	-3.140	19.18
Hampshire	14000	61.7%	74.04	46.74	74.722	-3.130	28.35
Herefordshire	986	69.5%	80.68	49.05	73.930	-3.600	14.42
Hertfordshire	6190	57.6%	80.02	49.52	75.452	-2.793	19.64
Humberside	230	27.0%	199.52	59.41	85.947	-3.380	54.76
Isle of Wight	421	43.0%	62.40	44.10	74.432	-3.852	26.23
Kent	21905	54.5%	85.91	48.74	75.958	-2.384	24.12
Lancashire	23093	50.7%	70.00	46.94	74.712	-3.200	32.92

Leicestershire	2788	45.3%	74.92	46.78	78.743	-2.379	24.56
Lincolnshire	4858	49.7%	69.54	47.02	75.739	-3.227	26.05
London	26975	45.5%	81.54	49.98	74.497	-2.642	45.20
Merseyside	7119	53.7%	70.14	47.01	76.112	-3.640	33.80
Middlesex	12269	55.5%	78.08	49.19	74.606	-2.706	29.45
Norfolk	5816	49.9%	78.61	48.13	76.733	-2.305	28.94
North Humberside	4908	47.3%	73.37	45.69	72.310	-3.674	33.20
North Yorkshire	4379	56.5%	79.74	49.34	76.353	-3.294	19.07
Northamptonshire	4878	49.2%	73.03	47.15	78.057	-3.029	21.35
Northumberland	2924	49.5%	50.79	43.52	77.983	-3.094	23.52
Nottinghamshire	9089	57.6%	70.63	45.64	75.684	-3.365	26.59
Oxfordshire	4117	56.3%	74.28	49.60	78.973	-3.117	16.49
Shropshire	2220	50.7%	80.51	49.71	74.451	-3.335	22.03
Somerset	11570	62.7%	81.15	49.07	78.421	-2.327	24.61
South Gloucestershire	1638	66.6%	85.45	50.00	74.322	-3.412	16.02
South Humberside	622	57.5%	46.11	39.73	72.432	-4.437	33.24
South Yorkshire	7275	47.8%	66.86	44.67	74.379	-3.514	32.63
Staffordshire	9556	47.5%	79.11	47.62	72.683	-2.994	29.64
Suffolk	3963	45.9%	95.66	50.82	79.283	-1.790	28.12
Surrey	8727	55.0%	73.79	48.12	75.523	-2.809	20.57
Tyne and Wear	8748	53.9%	73.64	46.32	76.039	-2.994	32.32
Warwickshire	4418	53.3%	83.62	49.89	77.517	-1.807	20.02
West Midlands	25374	49.4%	69.50	46.24	73.996	-3.510	40.15
West Sussex	8844	64.6%	77.44	49.36	75.527	-2.932	17.96
West Yorkshire	26407	59.9%	67.21	45.97	74.493	-3.171	30.17
Wiltshire	3469	58.5%	78.57	48.26	74.739	-3.214	21.36
Worcestershire	6857	54.5%	79.93	48.42	78.028	-2.691	20.73
mean			78.35	47.87	73.535	-2.945	26.42

Table 4: Top three primary schools in Wales by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP
1	2177682	Haberdashers Monmouth School*	Gwent	62.64	120.13	86.95	+0.05
2	2238278	St Illtyds RC Primary School	West Glamorgan	54.29	90.51	83.34	-1.38
3	1893583	Bryn Primary School	Dyfed	51.82	93.54	79.11	-1.27

\*private school. These data only for primary pupils.

Table 5: Top five secondary schools in Wales by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP
1	2128477	Ysgol Bryn Alyn	Clwyd	55.17	106.64	77.46	-3.83
2	1893869	The Maelor School	Clwyd	51.04	76.13	88.62	-3.67
3	2233693	Glyn Derw High School	South Glamorgan	48.39	92.66	75.22	-4.37
4	1963736		Dyfed	43.46	65.31	78.12	-3.17
5	1895439		Mid Glamorgan	41.46	46.37	70.63	-4.06



Table 6: Top five primary schools in Scotland by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP
1	2243944	Loretto RC PS	East Lothian	59.06	127.67	76.39	-0.35
2	1896785	Chapel of Garioch PS	Aberdeenshire	55.00	116.00	86.55	-1.18
3	2260604	Belhaven Hill School *	East Lothian	55.88	107.43	80.76	-0.22
4	2259591		Dumfriesshire	54.38	100.68	86.59	-1.13
5	2239058		Kincardineshire	52.79	108.71	85.76	+0.19

\*Private school. These data only for primary pupils.

Table 7: Top five secondary schools in Scotland by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidG P
1	2260604	Belhaven Hill School*	East Lothian	52.85	86.24	85.27	-2.77
2	1894975	Ellon Academy	Aberdeenshire	53.97	80.06	69.19	-2.51
3	2259591	Moffat Academy	Dumfriesshire	50.20	91.16	83.84	-3.19
4	2197856		Stirlingshire	44.68	36.74	69.33	+0.07
5	2006479		Lanarkshire	43.15	58.45	75.25	-3.88

\*Private school. These data only for secondary pupils.

Table 8: Top five primary schools in Northern Ireland by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP
1	2108301	Ampertaine PS	Co. London	72.92	189.62	88.01	-1.57
2	2235026	St Peters PS	Co. Tyrone	67.32	158.16	89.15	-0.60
3	1892521	St Olcans PS	Co. Antrim	65.33	132.70	89.79	-2.36
4	2235061		Co. Tyrone	63.56	133.49	92.32	-2.14
5	2235436		Co. London	61.91	106.52	88.51	-2.30

Table 9: Top five secondary schools in Northern Ireland by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP
1	2259586	Sullivan Upper School+	Co. Down	57.12	88.16	79.04	-2.75
2	2252017	Portora Royal School	Co. Ferman	50.95	102.33	91.11	-4.12
3	2252014	Bangor Grammar School+	Co. Down	50.40	74.55	89.02	-4.12
4	2260579		Co. Antrim	51.88	84.99	78.93	-4.15
5	2235152		Co. Down	50.12	80.98	81.95	-4.96

+grammar school.

Table 10: Top 40 primary schools in England by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP	Pupil Premium
1	3275654	Durston House School*	London	75.36	139.79	85.89	+0.27	0
2	1894764	Sandgate PS	Kent	73.37	178.22	82.64	-0.03	13
3	2248027	Redhill PS	Staffordshire	73.05	250.09	66.80	-1.44	8
4	2236630		Devon	69.21	173.17	89.41	-0.94	20
5	2239744		Hampshire	68.39	168.11	81.85	-0.68	9
6	2245046		Norfolk	67.23	150.23	79.85	-1.10	18
7	1892270		Middlesex	67.16	153.45	86.66	-0.43	37
8	2235885		Cumbria	66.52	167.05	82.14	-0.58	19
9	1893902		Staffordshire	67.29	140.28	63.91	-0.90	45
10	2233398		Buckingham shire	66.27	165.28	79.39	-0.56	16
11	2259873	* +	Buckingham shire	66.15	152.51	85.45	-0.67	0
12	2208923		Lincolnshire	65.46	161.03	89.47	-0.27	13
13	2231794		Somerset	65.40	163.13	76.84	-1.53	21
14	1893701		London	65.04	157.30	79.06	-0.78	55
15	2048242	* +	Buckingham shire	64.50	155.36	83.90	+0.17	0
16	1895076		Middlesex	64.44	161.82	70.11	-0.76	25
17	1962023		Tyne & Wear	64.13	141.77	73.44	-0.64	20
18	2245278		North Yorkshire	63.93	128.10	81.20	-1.30	16
19	1895780		Cumbria	63.80	158.78	86.04	-1.75	21
20	2236656		Devon	63.60	124.97	69.51	-0.24	30
21	2261020	*	Avon	63.50	143.67	85.39	-0.28	0
22	1950401		Cleveland	63.48	121.64	80.70	-1.28	43
23	2244153		Lancashire	63.27	140.56	79.63	-0.11	11
24	1978824		Cleveland	63.24	147.49	88.73	-0.98	28
25	2240654		Kent	63.12	160.21	76.82	-1.10	51
26	1896555		Kent	62.93	113.69	86.43	+0.10	16
27	2048540	*	Essex	62.65	134.06	85.61	-0.47	0
28	2236360		Devon	62.40	154.94	84.86	+0.01	0
29	2231932		Somerset	62.02	129.25	84.06	-0.70	10
30	2236617		Devon	61.77	124.05	78.96	-1.24	25
31	2235716		Co. Durham	60.99	150.73	81.50	-0.95	49
32	2029057		Devon	60.87	157.95	79.10	-0.34	19
33	2250232		West Yorkshire	60.80	130.61	81.84	-1.79	14
34	2235853		Cumbria	60.69	133.28	82.15	-0.93	19
35	1896676		Devon	60.60	129.21	84.96	-0.30	13
36	2232497		Berkshire	60.51	124.33	87.41	-0.72	29
37	2250952		Worcester shire	60.50	138.33	84.43	-1.47	11
38	2235887		Cumbria	60.23	114.43	73.49	-1.12	41

39	2259863	* **	Buckingham shire	60.03	140.65	86.74	-0.01	0
40	1895248		Somerset	60.02	133.89	84.79	-1.07	36

\*private school. \*\*boys only. +girls only. These data only for primary pupils.

Table 11: Top 19 secondary schools in England by outcomes and implementation

Rank	School Identifier	School name	School location	SGP	SS Gain	APC	ABL-MidGP	Pupil Premium
1	3385205	The Boulevard Academy	Humberside	85.45	390.58	85.75	-3.01	0
2	2232931	Q3 Academy	West Midlands	67.70	155.03	79.78	-4.62	19
3	2142013	Lancaster Royal Grammar School ** ++	Lancashire	65.18	97.07	85.54	-2.56	3
4	1892690		Lancashire	65.07	155.01	81.90	-3.55	42
5	2240748	**	Kent	64.56	174.32	76.90	-3.04	27
6	1894958	**	London	63.87	144.77	70.58	-3.07	0
7	2247407		Shropshire	63.77	173.33	69.98	-4.28	50
8	2236410		Devon	62.92	148.12	85.76	-2.49	12
9	1919269		Cornwall	62.59	129.67	87.62	-2.61	23
10	2119086		S. Yorkshire	62.62	140.61	77.67	-3.17	15
11	2248873	+ ++	Surrey	61.67	81.24	88.61	-2.36	4
12	1894690		Gloucester shire	61.92	150.06	77.58	-2.72	11
13	1893957		Hertford shire	61.59	136.81	85.11	-3.33	16
14	2237829	+	Essex	61.25	146.49	81.46	-2.69	13
15	1893291	** ++	Gloucester shire	60.82	126.81	85.95	-2.67	3
16	1893532		Surrey	60.72	134.83	79.59	-2.91	13
17	2237433	+ ++	Essex	60.66	88.12	83.85	-2.60	2
18	1919165	* +	Oxfordshire	60.59	143.82	83.24	-3.01	0
19	1892325		South Gloucester shire	60.33	128.37	76.14	-2.06	50

\*private school. \*\*boys only. + girls only. ++grammar school.